(2)

NorthWest Research Associates, Inc.
P.O. Box 3027 • Bellevue, WA 98009-3027

NWRA-CR-92-R090 18 December 1992

Final Report
Covering the Period 1 June 1992 through

NEW METHODS FOR NONLINEAR TRACKING AND NONLINEAR CHAOTIC SIGNAL PROCESSING

Prepared by
Jon A. Wright

Prepared for
Scientific Officer, Kam Ng
Office of Naval Research
800 Quincy Street
Arlington, VA 22217-5000

Contract No. N00014-92-C-0043
Attn: Code 122R4

300-120th Ave. N.E. • Suite 220, Building 7 • Bellevue, WA 98005 • (206) 453-8141 • FAX (206) 646-9123

NWRA-CR-92-R090

18 December 1992

Final Report
Covering the Period 1 June 1992 through

NEW METHODS FOR NONLINEAR TRACKING AND NONLINEAR CHAOTIC SIGNAL PROCESSING

Prepared by
Jon A. Wright

Prepared for
Scientific Officer, Kam Ng
Office of Naval Research
800 Quincy Street
Arlington, VA 22217-5000

Contract No. N00014-92-C-0043
Attn: Code 122R4

TABLE OF CONTENTS

1. INTRODUCTION

The classic problem in signal processing is to identify a low strength signal in a noisy background. Even more difficult problems are to identify several signals in the noisy background, to determine the direction of the sources, and to track one or more of these multiple sources. In the next section we will give an overview of the techniques that we have developed for nonlinear signal identification. In the third section we will provide the details. In the fourth section we will give an overview of our nonlinear tracking methods, and in the fifth section we will give the details of the design of an array suitable for use with broadband signals. In the sixth section we will present the details of our nonlinear tracking algorithm for following a maneuvering vehicle.

Classical, linear theory has been used extensively in solving these problems. However, if the signals have a broad frequency spectrum with no readily identifiable characteristics, standard spectral methods are inadequate. In Phase I, we have built upon some results of nonlinear dynamics to develop *nonlinear* algorithms to solve these problems, namely:

- classify sources of multiple signals in a noisy environment;
- determine the source direction if the receiver is an array;
- perform tracking of a maneuvering vehicle.

Imagine there are several sources of signals present in a noisy environment, and it is desired to classify the sources. Or suppose that you have been busy listening in a noisy environment and a new signal source suddenly appears which you wish to quickly identify. These are classic problems for which we developed a new generalized method in our Phase-I effort. Our method, which is based upon our research in nonlinear dynamics, specifically addresses signals that have a broad frequency spectrum with no easily identifiable features so that classical identification methods are ineffective. Our method is, to the best of our knowledge, completely new, and is based upon standard results in nonlinear dynamics and classical probability theory. It is particularly well suited to signals associated with low-dimensional chaotic dynamical systems, but it is by no means restricted to them.

We now give a brief overview of the procedures and assumptions involved in performing the classification. We assume that samples of signals have been previously acquired and are available in a library. The signal representing the noise background may be obtained while one is listening for other, more interesting signals, and it is treated the same as any other signal. Certain geometric structures (probability densities and their characteristic functions) associated with each signal then are constructed and a comparison is made with the same structure created from the newly received signal. The way in which this comparison is made can have a significant effect on performance. So far we have relied on either a standard χ^2 statistic for comparing two probability densities, or upon a simpler square of the difference of the densities summed over all signal strengths.

The geometric structures are defined so that the structure from the sum of two signals factorizes into the product of structures, one from each signal. We then compare with the

corresponding product of structures from our library and decide whether we have a correct identification. *This factorization property is crucial* as it permits a large library of signals to be compared with the incoming signal in essentially real time. The result is a determination that the received signal has a component that was made by a source of a unique type such as, for example, a particular helicopter model or a particular submarine model. For spectrally wide-band chaotic signals it is usually not possible to make such an identification by comparing either signals or their spectra.

One can imagine using this kind of processing in different ways. One scenario involves listening at a particular site with a nearly stationary (in time) noise background. Suddenly a new noise is present, and it is desired to identify the source. In this case we need only identify the new signal as we can include the observed background noise in our library. The searching of the library is then very rapid and an identification may be quickly performed. In other situations one may be confronted with a new environment and wish to identify all of the sources present.

In some situations the receiver is an array, and the signals from the individual elements can be separately phased (independent time lags) before combining the signals. In this situation it is possible to identify the direction of a particular source by varying the phase of the array. Because the signal is broadband, we are not looking for a null to determine the direction, but rather changes in the constructed geometric structures as the phases of the array elements are changed. To the best of our knowledge, this is a new concept. The densities constructed from noise-like signals were observed to have relatively little structure as the phase is changed, but the densities from other signals showed significant variation as a function of the array phase. It is also possible to monitor changes in the distance of the source as the amplitude of the received signal varies. A new algorithm that we have developed determines the velocity of the source by a nonlinear "doppler shift" calculation. This determination is sensitive to the background noise; consequently, its practical implementation will have to be delayed until the nonlinear signal-processing algorithms are refined.

In conjunction with the development of our nonlinear identification algorithms, we have also derived a nonlinear tracking algorithm. It also turns out to be suitable for tracking maneuvering vehicles in a conventional signal-processing environment.

In Phase I we performed the following tasks:

- We constructed twenty signals, including several types of broadband noise, outputs of chaotic electronic circuits, and samples of time series generated by systems of ordinary differential equations with chaotic solutions. Samples of the signals were placed in a "library." Additional signals from the same systems were generated and constitute the "test" set. The signals in the test set were not identical to any in the library, but they were from the same dynamical systems. In addition, we forced some of the noise signals to have the same spectrum as some of the chaotic signals. This was accomplished by Fourier transforming the signal and randomizing the phases of the Fourier components and then inverting the Fourier transform. This insures that standard spectral-identification methods could not separate the sources.

Two signals from the test set were added together with randomly selected amplitudes to form a simulated received signal. We then attempted to identify both the amplitudes and the type of signals by comparing certain geometric structures constructed from the signal and from the library. Provided the amplitudes of the test signals were not too different, we were able to identify both signals in all cases. The range of amplitudes permitted depended upon the similarity of the signals, with a factor of roughly ten being a limit for the modest amount of processing that we did. We believe that as the algorithms are refined, substantial improvements in the range of amplitudes for which signals can be separated will be possible.

- We also combined three test signals together for a few examples and were able to classify the components.

- We next considered an example where the receiver was a multi-component array. The incoming signals from the receivers in the array were added with a phase difference which was chosen to select a particular direction. Because the signals are broadband, there is no "null" as a function of the relative phases of the array elements. However, the constructed densities have a strong dependence on the phase, which can therefore serve both as a direction identifier and as an additional classification mechanism. In examples of arrays that were typical of ones that could be deployed in the ocean, we were able to determine the direction and identify the source of signals that were 20db weaker than other signals with similar spectral characteristics with only a modest amount of processing.

The techniques that we developed are new, and there are a wide variety of possible refinements. Because the processing is nonlinear, there are more variations available than for conventional linear signal processing, even when we restrict the investigation to the types of algorithms that we propose. There are no real differences in the principles involved in the various implementations, but there are many different ways to optimize the probability of correct classification. Unfortunately it is very difficult to prove general theorems about what is optimum for nonlinear systems. Thus we expect that, for the next few years at least, the non-linear signal processing community will have to rely to a large extent on empirical evidence.

2. OVERVIEW OF OUR SIGNAL CLASSIFICATION METHOD

We now discuss some of the important variables and some of the freedom available in the details of our algorithms. Usually our signal is a time series from a single detector. In nonlinear dynamics we are often interested in quantities that might have been observed if more detectors were used. In order to get some additional information, we construct additional surrogate signals. There is a certain arbitrariness in this procedure which leads to a wealth of simple variations in processing algorithms. For example, the most common way of constructing a second signal is to start with the original signal, $S(t)$, and construct another signal, $S_1(t) = S(t + T)$ where T is a time delay. The two signals, $S(t)$ and $S_1(t)$ can be used to construct a two-dimensional phase space. Other signals can be constructed by using other time lags, such as $S_2(t) = S(t + 2T)$. For dynamical systems with low dimensional

attractors, it is possible to construct a phase space that has geometric properties related to the actual phase-space geometry. This technique of phase-space reconstruction has been used extensively to obtain information about dynamical systems. It underlies our procedures, but we have found that it can be extended in several ways for purposes of signal recognition.

In most nonlinear dynamics applications, one considers a fixed time delay, T, and for those applications nothing is gained by varying the time delay, except perhaps one choice works slightly better than another. For our applications it is useful to consider several time delays simultaneously. Of course, this is not a new concept in linear signal processing since the auto-correlation function considers all time delays. However, this idea has not previously been used for examining phase-space density functions. We found that making separate comparisons for several different time lags improved the decision-making algorithm. As one example, an alternative way of constructing a second signal is to consider the signal $S_{sum}(t) = S(t) + S(t+T)$. Although this is an apparently trivial change, it turns out to be a crucial ingredient when the receiver is an array, and it provides a significant improvement in our classification algorithm for single receivers. The density constructed from such a signal combination is a slice through a two-dimensional density constructed from S_1 and S_2, but it usually has more structure than the density constructed from S_1 alone and is hence more useful.

The point of the above two paragraphs is that there are a variety of simple modifications that can be made to our basic method that will improve the results. As we have indicated above, there are several variations in the numbers and types of phase-space densities that can be used for identification. The principal ones that we have tested are as follows:

- Two-dimensional phase space with a single, fixed delay time;
- Two-dimensional phase space with one or two additional delay times;
- One-dimensional phase space using $S_{sum}(t)$ and many delay times;
- Three-dimensional sampled phase space with a fixed delay time.

Most of our tests have been performed using a two-dimensional phase space with a fixed delay time, T, and it is within that framework that we declare Phase I to be a success. In order to improve the discrimination, we first considered a second time lag and that was sufficient to remove some accidental ambiguities. It is possible, and in principle desirable, to work in a three-dimensional or higher phase space. Unfortunately, the computational time increases dramatically. Rather than reconstruct the entire three-dimensional density, it is possible to sample randomly the density (actually the Fourier transform of the density) at many points. This allows one to enjoy many of the benefits of a full reconstruction without the full computational effort. We performed some limited testing using this procedure in Phase I and it appears to be successful. If several signals are present, this technique might provide a better signal to noise ratio for the same computational effort. We have not yet made comparisons with the use of two-dimensional densities in terms of efficiency and accuracy.

We also did some calculations with one-dimensional densities using the signal $S_{sum}(t) = S(t) + S(t+T)$ described above. The disadvantages of a density constructed from just $S(t)$ is that it is difficult to obtain good discrimination because of the relatively few features available for discrimination, and that drawback is related to the relatively little information contained in that density. However, adding a time-lagged signal increases the structure and considering many different time lags greatly increases the identification capability. A significant advantage of one-dimensional densities is that relatively little computer resources are required. We have tested this concept on a few signals. At the very least it is successful enough to serve as a quick "rejection filter," thus reducing the number of cases requiring more extensive examination.

As we mentioned above, these combinations and their generalizations also play an important role when the receiver is an array. In the plane-waves approximation, the best one can hope to do is to translate the phase information into a bearing angle. The problem is thus to design an array whose beam pattern has little or no frequency dependence. In particular, we must make sure that the main lobe width, peak response, sidelobe level, and distance between the main lobe and the sidelobe plateau (which takes the place, in our approach, of the grating sidelobes which appear generically in the case of single frequency arrays) all remain constant across the design frequency band. To do this, we use the Poisson summation formula and the method of stationary phase to relate beam-pattern properties to array characteristics. Using this relation, we then translate all the listed beam-pattern requirements into practical requirements on sensor location and array amplitude shading. We end up with a beamforming approach that is significantly more efficient than a classical uniformly spaced array. Specifically, while uniform spacing produces an array in which the total number of elements would be proportional to the ratio of the highest to lowest design frequencies, the method we have used allows one to achieve all the beam-pattern objectives with a total number of elements that grows like the logarithm of the ratio of highest to lowest frequencies.

In summary, in Phase I we have developed new, nonlinear algorithms for detecting one or more spectrally broadband signals in a noisy background. The primary goal of verifying that signals can be identified/classified by the proposed technique was achieved. In addition, the fact that there are several variations of constructing densities that have been successful in identification seem to enhance greatly the ideas' usefulness. The algorithms use a new and innovative approach that can be used in near real time. We have successfully used the algorithms to classify sources and, with an array of receivers, we can identify the direction as well as decrease the contribution of other sources.

3. SIGNAL CLASSIFICATION

The crucial ingredients for a correct classification of signals are their probability densities. First, we explain two ways of constructing the densities, and then we will show how to identify the components of a signal made up of several pieces. Examples will be given showing how the method works, and some important refinements will be presented.

The first step is to normalize all signals. The mean is removed and the root-mean-squared signal is normalized to unity. It is essential to have a uniform normalization since

we have no a priori knowledge of the relative amplitudes of the components of a signal. The normalization and mean may be useful themselves, such as for determining the change in distance of a moving vehicle, but we don't use them directly for most signal identification.

The simplest density is just a histogram of the amplitude of the incoming signal. If $S(t)$ is sampled and has the values $S_1, S_2, S_3 \ldots$ for a total of N values, we just count the number of times S is between x and $x + dx$ and call that $N \cdot \rho(x)dx$. Formally,

$$\rho(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \delta(x - S(t))dt \tag{1}$$

where $\delta(x)$ is the delta function. This density usually doesn't have enough structure to be a useful identifier, although with a modification to be described below it can be a useful tool. A second density, now two dimensional, can be constructed by introducing a second coordinate $S(t + T)$. Then

$$\rho(x, y) = \lim_{T \to \infty} \int_0^T \delta(x - S(t))\delta(y - S(t + T))dt. \tag{2}$$

The practical application of these formulae requires replacing the delta function by a function of finite width. We have chosen to use

$$\delta(f) \to \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-f^2}{2\sigma^2}}. \tag{3}$$

Other substitutions are possible.

The interpretation of $\rho(x, y)$ was discussed in the previous section. A third coordinate can be introduced and a three-dimensional density defined. The procedure could, in principle, be continued to whatever dimension one wished. There are several reasons for truncating at a low dimension:

- Computational requirements become prohibitive in high dimensions (greater than three)
- The length of the signal may not be long enough to justify many dimensions, i.e., regions of phase space will be empty or inadequately sampled.
- For low-dimensional dynamical systems the higher dimensional probability densities are certainties in terms of lower dimensional ones.
- A density in a low dimension is a projection of a density in a higher dimension. The low-dimensional density may contain sufficient features to serve as a classifier for practical purposes.

It is possible to sample a high dimension without computing the entire density. This procedure will be discussed below. For dynamical systems with low-dimensional attractors,

strange or otherwise, there are theorems that relate the number of time lags needed to the dimensionality of the system. Consequently, if only low-dimensional dynamical systems were the sources of the signal, we need consider only lower dimensional probability distributions. Noise tends to be very high dimensional, but it usually has little structure anyway so not much is lost by projecting it onto a lower dimensional space.

We have elected to do our calculations mostly in a two dimensional space. As we have noted, the various time lags are parameters of the density functions. To work with only one time lag, T, is equivalent to considering the auto-correlation function

$$C(T) = < S(t)S(t+T) > \tag{4}$$

at only one value of T. Of course one value of $C(T)$ provides very little useful information. For example, the determination of the spectrum requires a wide range of values of T (in principle all T values, in practice a limited range). Thus even when we restrict our calculations to two dimensions we can construct a continuum of densities depending on T. In practice we have sampled a range of T values.

The next step in the procedure is to understand the relationship between a density of a signal and the density of its components. To this end we define the characteristic function of the density which is just the Fourier transform,

$$\hat{\rho}(k) = \int e^{ikx}\rho(x)dx = \lim_{T\to\infty}\frac{1}{T}\int_0^T e^{ikS(t)}dt. \tag{5}$$

This has the property that if $S(t) = S_1(t) + S_2(t)$ and

$$\hat{\rho}_1(k) = \frac{1}{T}\int_0^T e^{ikS_1(t)}dt \tag{6}$$

$$\hat{\rho}_2(k) = \frac{1}{T}\int_0^T e^{ikS_2(t)}dt \tag{7}$$

then $\hat{\rho}(k) = \hat{\rho}_1(k)\hat{\rho}_2(k)$. This is a well-known property of characteristic functions [5] and just requires that S_1 and S_2 are generated by independent, stationary processes.

We emphasize that this relationship is crucial to the possibility of signal classification if multiple signals are present. If three signals are present the result is

$$\widehat{\rho}(k) = \hat{\rho}_1(k)\hat{\rho}_2(k)\hat{\rho}_3(k). \tag{8}$$

This expression would be correct for general signals S_1 and S_2, but since we normalize signals, the amplitude must enter the relationship between the $\widehat{\rho}$s in Eq. 8. Suppose

$$S(t) = \alpha S_1(t) + \beta S_2(t) \tag{9}$$

where all signals have zero mean and are properly normalized. Then $\alpha^2 + \beta^2 = 1$, and the relationship is actually

$$\widehat{\rho}(k) = \hat{\rho}_1(\alpha k)\hat{\rho}_2(\beta k) \quad . \tag{10}$$

Since α and β aren't known, this requires that they be found as part of the procedure. The relationship is the same in higher dimensions except that k is now a vector.

The density can be constructed either by constructing $\rho(x)$ or $\widehat{\rho}(k)$ from the above definitions. It is usually more efficient to calculate $\widehat{\rho}(k)$ on a mesh and then interpolate to get $\widehat{\rho}(\alpha k)$, rather than calculate $\widehat{\rho}(\alpha k)$ directly. Note that $\rho(x)$ will always vanish for x outside some region, since signals can't have arbitrarily large amplitudes.

Once the appropriate densities are constructed, the next step is to test hypotheses about which signals make up S. The procedure that we have used thus far to provide a figure of merit is to test the quantity

$$\text{Error} = \sum_k \mid \widehat{\rho}(k) - \hat{\rho}_1(\alpha k)\hat{\rho}_2(\beta k) \mid^2 \tag{11}$$

or the corresponding Fourier transform as a function of x for all signals, S_1 and S_2, for a range of values of α. If one of the components is known then one need only search for a second signal. This is the case if a new signal suddenly appears. The background signal up to that time is S_1 and the new signal is S_2. Since we have been listening to S_1, we can construct the densities for it and only candidates for S_2 need to be tested. This also illustrates that one of the signals, S_1 in this case, can be arbitrary noise. If there are three signals present, all unknown, one has to test many more hypotheses.

We have found that a special class of one-dimensional densities, to be described below, can be used as a rejection filter. This is an important point since they require very little computational resources.

We now present examples to illustrate the method. In Figures 1 and 2, several examples of signals are shown. Two of the signals are from electronic circuits (provided by NRL) and the rest are synthetic signals generated from ordinary differential equations with chaotic attractors. Figures 3 and 4 show the power spectrum of the same signals.

The basic idea of the scheme we propose for classifying signals is to construct probability densities for signals of interest and to make comparisons by pattern matching with the corresponding density for a signal that one wishes to identify. There are an infinite number of possible densities, but it is sufficient to compare only a few. Before proceeding to the technical details we give a pictorial example. In Figure 1 there are three density plots. The first, (a), is the density constructed from a signal to be classified and the other two, (b)

and (c), are proposed candidates taken from a library. In this case it is easy to see that (c) matches (a) and that (b) does not. A numerical calculation of the differences confirms this. This particular density has a simple interpretation. It is the probability that a measurement of a signal, $S(t)$, at time, t, and a subsequent measurement a time T later will have the values x and y respectively. Or more properly, $\rho(x,y)dxdy$ is the probability for the signal $S(t)$ to be in the range x to $x+dx$ and the signal $S(t+T)$ to be in the range y to $y+dy$. $\rho(x,y)$ is the vertical axis labeled density. The densities will, of course, be different for different values of T. The second moment of the probability density, $\rho(x,y)$, is the autocorrelation function, i.e., $C(T) =< xy >$.

Such a simple comparison does not always provide a conclusive test, particularly when multiple signals and noise are present. In that case one can compare the patterns at several different values of T or construct other related densities. There are many (infinite) densities that can be constructed and it must be determined which are more useful for a particular application.

This example illustrates some of the flexibility available in our algorithms for signal discrimination, and it also serves to illustrate some of the problems that need to be solved in order to develop a useful system. There is flexibility in the choice of dimension of probability density, there is the question of selection of the time lag, T, and whether to use many different values of T and, if so, how many. There is the possibility of using an array of receivers to enhance recognition. Since the sources have a broad frequency spectrum, the question arises as to the most effective way of deploying an array and what are the trade-offs. One also needs an automated decision making process for comparing densities and some way to measure confidence. All of the suggestions for enhancing performance were tested successfully in Phase I, although no information about optimizing performance was obtained.

In order to test our method when multiple signals are present, we form combinations of two signals to give a third,

$$S(t) = aP(t) + bQ(t) \tag{12}$$

The signals P and Q are to be identified as well as their relative amplitudes. Since we measure $S(t)$ we know its level so that we know the value of the sum of the squares of a and b. For the particular example that we show here, $a = 0.48$ and $b = 0.87$.

We have been using a χ^2 statistic as a figure of merit, which is a slight modification of Eq. 11 [5] and the appropriate function of the densities of the signals is compared. In this case we constructed a two-dimensional density by using a signal and a time-lagged signal for the two components.

The criterion that we use is exact in the limit of an infinitely long signal, but in the present case a relatively short signal was used. This means that even when the correct signals with the correct amplitudes are compared there will be some residual statistical error. In the lower right graph of Figure 6, χ^2 is shown versus the relative strength parameter, α. There are two curves that are nearly identical and a third one that has a larger χ^2. One of

the two lower curves corresponds to the exact comparison and gives us an estimate of the statistical uncertainty that places a lower limit on χ^2. The second curve was computed by taking a different sample of signals 4 and 6 and comparing various combinations of them with the densities of the original signal. We see that the value of α is clearly about 0.5, which is correct.

The residual error is due to the inherent lack of statistics. The upper curve corresponds to the hypothesis that one of the signals was correct and the other incorrect. Note that even in this case our algorithm correctly identifies the amount of the signal that was the right candidate. We then considered two different signals, neither of which corresponded to the original choice, namely signals 1 plus 5 (see lower left of Figure 6). In this case χ^2 was relatively flat and well above the levels shown on the graph, thus affirming that neither signal was present in the signal to be identified.

The amount of data used is of course important. The computational effort is proportional to the length of the time the signal is observed, and there is the logistic problem of not being able to collect an infinite amount of data in a finite time. For the examples shown, we sampled about ten times per oscillation and included a total of 10,000 points. For typical machinery oscillations of 1,000 rpm, this is about a minute's worth of data.

Depending upon the types of signals of interest, one may get by with less total samples. We tried many different combinations of signals, even including noise in the signal to be identified, and all of our results were comparable to the example presented above. We were even able to recognize that a signal was made up of two similar signals so that if there are two sources of the same signal we could recognize that fact.

It is important to be able to identify a weak signal combined with a strong signal. The limitations on this are not known at present and will undoubtedly be dependent upon refinements of the algorithm. Qualitatively we have no problem separating signals with a ratio of energies of ten to one. Sometimes we can resolve power levels of a hundred to one, and we expect further improvement will be possible.

We have done some limited testing with a signal made up of three signals. In the first example one signal was broad-band noise (signal 7) and the two other were chaotic signals. The energy in each signal was the same. The correct combination was easily selected out. We then tried a combination with the energies in the ratio of 9 to 4 to 1, with the 9 being the broad-band noise. There did not seem to be any problem identifying the two signals even in the presence of noise. There was one example of an alternative signal giving as good a statistic as one of the correct signals. In order to resolve this discrepancy we performed a separate test which completely resolved the uncertainty. It is encouraging that almost all incorrect combinations were easily rejected in the two-dimensional phase space, as this means that only a few candidates might have to be further processed, as we did in the above example in order to get better rejection of incorrect hypotheses.

In order to give the reader a visual impression of the densities we illustrate with a combination of two signals, such that the weak signal has 40% of the amplitude of the first signal. We then consider three candidate signals for the weak signal. Figure 7a shows the

density $\rho(x, y)$ of the incoming composite signal with $x = S(t)$ and $y = S(t+4)$. The density obtained by taking the convolution of the two correct individual signal densities is shown in Figure 7b. Note that density in Figure 7b is smoother than Figure 7a. This is due to the limited statistics. If the signals were infinitely long, the two densities would be identical. Incorrect hypotheses for the weak signal combined with a correct guess for the strong signal are shown in Figures 7c and 7d. Note that visually we can distinguish them, and a numerical test also concludes which is the correct density.

We now turn to a promising method of performing some classification via one-dimensional densities. Define a new signal, $R(t) = S(t) + S(t + T)$. Construct the density associated with this signal.

$$\rho(x; T) = \frac{1}{T_o} \int_0^{T_o} \delta(x - R(t))dt. \tag{13}$$

As usual we normalize $R(t)$ such that $rms(R^2) = 1$. The factor involved in the normalization is just the correlation function, that is, $rms(R^2) = 2(1 + C(T))$ before normalizing R but after normalizing S. The density $\rho(x; T)$ has considerable structure and the structure changes as T changes. In Figure 8, we present examples for one particular signal. The density is shown on vertical axis and the signal strength, x, is on the horizontal axis. The value of T is shown at the top of each graph. For reference, a typical oscillation time scale is around ten units. The densities for a second signal are shown in Figure 9. Note that they are quite distinct, particularly if several different T values are examined. The density with $T = 0$ corresponds to the usual density. Note that it is relatively featureless. One way of constructing noise with a frequency spectrum comparable to a signal is to Fourier transform the signal and make the phases random with each one independent of the others. Then Fourier transform back and the resulting signal will have the same spectrum. The result of such an operation is shown in Figure 10 a-f. Notice that there is no structure and no dependence on T. Figure 10g is for $S(t) = \cos(t)$. There is no dependence of $\rho(x)$ on T for a simple sinusoidal signal.

We have done some limited testing of comparing densities in one dimension as a function of T. It is certainly much faster than computing in two dimensions and serves as an effective rejection filter. Properly used, it might be possible to do all or nearly all calculations with one-dimensional densities. One can extend the idea and define

$$Q(t) = S(t) + S(t + T) + S(t + 2T). \tag{14}$$

The density associated with $Q(t)$ has yet a different structure. One could also use a different T, say T_1 instead of $2T$. All of these concepts need to be examined. We now know that they all work to some extent and the goal is to refine and optimize their use.

In order to illustrate the possibility of discrimination with one-dimensional densities, we chose two signals and added them together with equal amplitudes. In Figure 11 the solid line is the density of the summed signal. The dashed line is the fit obtained as usual by

using $\widehat{\rho}(k) = \hat{\rho}_1(k)\widehat{\rho}_2(k)$. The curves would exactly coincide if the length of the signal were sufficiently long. The difference in the two curves provides a measure of what is the "best" that one can do if one tries candidate signals to make up the received signal. In this case the most stringent test was to try a signal with the same power spectrum as one of the original signals, but with randomized Fourier phases so that it is noise. The result is shown in Figure 12. Although the fit looks good, they are to be compared with those in Figure 11. There is clearly a difference for all values of T and one could easily infer the correct signal even with this one-dimensional construction.

The one-dimensional densities are just simple projections of two-dimensional or higher densities. However there are alternative projections that are also simple. For instance, for the Fourier transform in two dimensions, $\widehat{\rho}(k)$, one can fix the magnitude of k and vary the polar angle. We have done some limited testing of this idea and the ability to classify seemed comparable to the one-dimensional densities described immediately above. The reason for considering these alternatives is that it is desirable to use a projection that has considerable structure, and the region around $k = 0$ is generally pretty structureless. Ordinary projections have to use that region whereas this projection is specifically devised to avoid it. It also has the feature that it is simple to calculate and the interpolations required because of the unknown signal strength are also simple.

It is possible to extend the definition comparable to $R(t)$ or $Q(t)$ to two dimensions. We have done that in a limited way and the comparisons seem to work better; however, not much testing has yet been done.

An important new tool that we have been developing is the use of an array of receivers. We use the array first to get the direction of a particular source and then to analyze the source. We made a model of an array typical of a possible underwater array. The particular test we ran was for one hundred elements, spaced a few meters apart. The details of designing an appropriate array are presented in the next section.

Both equal spacing and logarithmic spacing were tried. Two sources a long distance away with an angle of 45^o separating them were simulated. In one example we made source A ten times as strong as B (20db separation). We were easily able to identify B with a single two-dimensional comparison or a few one-dimensional comparisons. We could not do this without the array. In order to demonstrate the above comments we show the results with and without an array. Without an array it is just the addition of two signals with one signal being ten times larger in amplitude. In Figure 13a we show the density the sum of the two signals,

$$S(t) = S_1 + 10S_2. \tag{15}$$

In Figure 13b we show the density obtained by the usual multiplication of $\widehat{\rho}_1$ and $\widehat{\rho}_2$. In Figure 13c we show an alternative candidate for S_1. Clearly it is difficult for the eye to see the difference between Figure 13b and Figure 13c. It is also numerically difficult if only one time lag is used. In Figure 13d we show the density for an alternative signal for S_2. This time we can clearly see the difference. In Figure 14 we show the results of using an array. This time we are clearly able to identify correctly the weak signal. Note that since the signal

is broadband, we can't use the array to select a particular frequency in a given direction as one might do with harmonic signals. This is a very important result since we don't have to search for pairs of signals, etc. but only need to match a single signal, as the other received signals were greatly reduced in strength.

4. OVERVIEW OF BROAD BAND ACOUSTIC ARRAY DESIGN

We studied the problem of localizing and tracking the source of a received underwater acoustic signal, given that the latter is more complex than a simple CW tone, and that it has propagated from source to receiver through a non-linear medium. The approach we had to take is a dual one: it depends on how well one can represent the detected signal. Indeed, if the signal produced by the source is truly chaotic, i.e. if it is has chaotic dynamics, then, while we have an approach that can identify the dynamical system in question, one cannot hope to retrieve the actual time-series of the signal with any kind of accuracy. On the other hand, if the signal produced by the source is merely complex, i.e. if it is more complicated than a (linearly-produced) time-harmonic tone yet has Hamiltonian dynamics or at least non-chaotic ones, then one can indeed hope to identify a template time-series associated to it, perhaps parametrized by some of the state variables, then estimate these parameters along with the other state variables associated to the source.

In either case, the signal produced by the source gets convolved by the Green's function of the medium before being measured by the receivers. Because the source is moving in time, and because the dependence of the Green's function on range is very much non-linear, the received signal is quite complex in both cases. Yet one would expect that if this complexity is well-understood and optimally used, one could infer from it the location of the source as a function of time. In turn, it is natural to expect that this localization and tracking should help improve the detection and estimation of the various source parameters.

In the non-chaotic case, one can try to model the template associated to the source signal as accurately as possible: indeed, it will depend on the source coordinates (which vary with time), on its (time-varying) velocity, and, possibly, on some other (non-time-varying) parameters that can distinguish one underwater source from another. Once the dynamics of these state variables are modeled by mathematical equations (with deterministic and stochastic components), and once the dependence of the received signal on these variables is also modeled mathematically (i.e., essentially, once the medium is satisfactorily represented and its Green's function computed), we reduce the problem to a classical optimal-filtering problem, albeit a non-linear one. The approach we take in this case is to derive and solve the equations for the conditional density function. This is the probability density function of the state variables describing the problem, conditioned on the signal received. Computing this p.d.f. is essentially the equivalent of implementing an infinite-dimensional Kalman filter to provide the best estimate of the state variables in this non-linear problem.

In the truly chaotic case, one cannot expect to single out an actual signal template. Yet the approach we took to the identification problem does associate a characteristic density function to our source, one shared by all the diverging actual signals that could be produced by this single dynamical system that has been identified. As this density function scales linearly with amplitude, we can also estimate the instantaneous (real) amplitude of our

signal. In order to develop a tracking method in this case, the challenge is two-fold: use plane-waves (rather than the detailed Green's function) to make the best use of what little phase information one has, then try to refine this angle-tracking by making optimal use of the sequence of estimates of the instantaneous real amplitude information we obtain using the density function that characterized the signal. Our tracking approach in the chaotic situation is thus somewhat weaker than in the previous case: instead of making optimal use of the complete complex Green's function, we first account for all phase information using a plain-wave approximation, then account for the varying amplitude using the real propagation loss function (the magnitude of the Green's function).

In the next section, we describe our filtering approach in the chaotic case in more detail, and present simulated results. We then describe our approach in the non-chaotic-yet-nonlinear case and present simulated results in that case.

5. BROAD BAND ARRAYS

As explained in the previous section, we first seek to extract all phase information from our signal before processing it through our dynamical system identifier, followed by our instantaneous amplitude estimator. The resulting amplitude estimate is then used along with the phase information as the latest data for our tracking algorithm. We now describe the individual pieces of this method.

In the plane-waves approximation, the best we can hope to do is to translate the phase information into a bearing angle. The problem is thus to design an array whose beam pattern has little or no frequency dependence. In particular, we must make sure that the main lobe width, side lobe level, and distance between the main lobe and the side lobe plateau (which takes the place, in our approach, of the grating sidelobes which appear generically in the case of single frequency arrays), as well as the peak response, all remain constant across the design frequency band. To do this, we use the Poisson summation formula and the method of stationary phase to relate beam-pattern properties to array characteristics. Using this relation, we then translate all the listed beam-pattern requirements into practical requirements on hydrophone location and array amplitude shading. This approach is similar to the one proposed by Ishimaru in [3] [4], and differs in that, while Ishimaru et al only verified that a particular shading function gives a good sidelobe behavior, we use the approach to systematically derive the array requirements in order to achieve our global broadband objectives. Our generalization of his method to the systematic broadband case here follows the same line as our generalization of his method to time-domain pulse trains ([2]).

Let us begin with some design considerations. The problem with using a classical uniform element spacing comes from the two facts that

- in order to reduce the side lobe levels across the band, the spacing would have to be proportional to the wavelength corresponding to the highest design frequency, while
- in order to keep the width of the main lobe constant over the design band, the length of the array would have to be proportional to the wavelength corresponding to the lowest design frequency.

In other words, with uniform spacing, the total number of elements would be proportional to the ratio of highest to lowest design frequencies. With our design, it turns out that the number of elements will grow like the logarithm of this ratio, hence should prove substantially more efficient.

The main difficulty in using unequal spacings is that the beam pattern function for such an array is unlikely to have a closed form expression. This makes it difficult to relate array beam pattern properties to array specifications (i.e. spacing and amplitude shading). That is where the Poisson summation formula comes in.

Suppose we have $N+1$ elements, to be placed along the x-axis at positions x_n ($0 \leq n \leq N$), and to be shaded by the (frequency-dependent) weights $w_n(\lambda)$, λ a wavelength within the band of interest. As a function R of α = sin (angle of the plane wave arrival) $-$ sin (array beam angle), the response of the array will be

$$R(\alpha, \lambda) = \sum_{n=0}^{N} w_n(\lambda) e^{2\pi i \alpha x_n/\lambda} \tag{16}$$

There are many ways of extending x_n and $w_n(\lambda)$ to functions of a real argument n such that

- both functions are continuous at all the integers,
- x is twice differentiable and strictly increasing on the interval $[0, N]$,
- w vanishes for n not in $]-\epsilon, N+\epsilon[$ (where ϵ is a small positive number fixed once and for all).

Choose such extensions $x(n)$, $w(n, \lambda)$. We can then use the Poisson summation formula on R to get

$$R(\alpha, \lambda) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} w(n, \lambda) e^{2\pi i \alpha x(n)/\lambda} e^{-2\pi i m n}\, dn \tag{17}$$

For simplicity, let us assume that the element spacings are increasing (with n). Mathematically, we are assuming that the derivative x' is an increasing function, i.e. that the second derivative $x''(n) > 0$ for n in the interval $[0, N]$. Following [3] [4], let us use the method of stationary phase to study the expression (17). This method predicts that, asymptotically, (i.e. for λ relatively small), the behavior of R for α close to 0 will be adequately described by that term in the right hand side of (17) which corresponds to $m = 0$ (i.e. the mainlobe is described by the $m = 0$-term), while the m^{th} grating sidelobe will be described by the behavior of the integrals near that $n = n_m$ which gives a stationary phase in the integrand, i.e. for which

$$x'(n) = \frac{m\lambda}{\alpha} \tag{18}$$

Since x'' was assumed positive, x' is strictly increasing, so (18) will have at most one solution for each m.

Let us begin by examining the mainlobe-to-sidelobe separation problem. What is the smallest positive value of α for which (18) has a solution, i.e. where does the first sidelobe appear? For α sufficiently near 0, the right hand side of (18) is very large, outside the finite range $x'([0,N])$ of x'. As α increases, the right hand side will decrease accordingly. The first solution thus occurs when $\alpha = m\lambda/x'(N)$. Hence, at the lowest design frequency, where we can and will make the reasonable assumption that the whole array is used, the first ($m = 1$) sidelobe appears when $\alpha = \alpha_0 = \lambda_0/x'(n)$, where λ_0 is the wavelength corresponding to the lower design frequency, and $x'(N)$ is, by hypothesis, the largest element spacing. In order to meet our requirements, we would like to keep this main-to-side-lobe separation constant across the design band. The obstacle is that, if $\lambda < \lambda_0$, (18) will have solutions with $\alpha < \alpha_0$. The only way to remove this hurdle is to make sure that the corresponding shading weight equals zero, for all these potential solutions. So we have to set

$$w(n,\lambda) = 0 \text{ for all } n > x'^{-1}(x'(N)\lambda/\lambda_0) \tag{19}$$

and keep it positive otherwise. This equation has a nice interpretation: it says that at any given frequency in the design band, elements spaced farther than $x'(N)/\lambda_0$ wavelengths apart should not be used.

Let us now examine the sidelobe level problem. As in [1], (17) implies that the level of the m^{th} grating sidelobe is approximately

$$|\int_{-\infty}^{\infty} w(n,\lambda) e^{2\pi i(\alpha x(n)/\lambda - mn)} dn|^2 \tag{20}$$

which, by stationary phase, is itself approximately equal to $\lambda w(n_m,\lambda)^2/(\alpha x''(n_m))$, with $n_m = x'^{-1}(m\lambda/\alpha)$ as in (18). That is, the level of the m^{th} sidelobe is approximately

$$\frac{1}{m} x'(n_m) \frac{w(n_m,\lambda)^2}{x''(n_m)} \tag{21}$$

As we determined in the previous paragraph, the first (and largest) sidelobe occurs at $\alpha = \lambda/x'(N) > \alpha_0 = \lambda_0/x'(N)$, i.e. for α in the interval $[\lambda_0/x'(N)\ ,\ 2]$, for which n_1 must correspondingly lie in the interval $[x'^{-1}(\lambda/2), x'^{-1}(x'(n)\lambda/\lambda_0)]$. For this sidelobe level to be independent of λ, we have to require that

$$w(n,\lambda)^2 = constant \frac{x''(n)}{x'(n)} \text{ if } x'^{-1}(\lambda/2) < n < x'^{-1}(x'(N)\lambda/2) \tag{22}$$

over all λ in the design band.

Let us now turn to the mainlobe level problem. As in [1][3], the array response for α near 0 is approximately given by the 0^{th} term in (17), namely

$$\int_{-\infty}^{\infty} w(n,\lambda) e^{2\pi i \alpha x(n)/\lambda}\, dn \tag{23}$$

Choosing $t = x(n)$ as a new variable as in [2], this can be rewritten as

$$\int_0^{L_0} w(x^{-1}(t),\lambda) e^{2\pi i \alpha t/\lambda} \frac{dt}{x'(x^{-1}(t))} \tag{24}$$

where we have made the reasonable assumption that the array extends from $0 = x(0)$ to $x(N) = L_0$. There are several ways to make sure that this integral is independent of λ, at least for α small. One simple way is to write $L(\lambda)$ for the length of that part of the array used at wavelength λ, i.e. to assume that $w(n, \lambda)$ vanishes exactly when $n > x^{-1}(L(\lambda))$, so that the integral can be rewritten as

$$\int_0^{L(\lambda)} \frac{w(x^{-1}(t), \lambda)}{x'(x^{-1}(t))} e^{2\pi i \alpha t/\lambda} \, dt \tag{25}$$

then to assume that the integrand is as spatially uniform as possible, i.e. depends only on λ. We thus assume that $w(x^{-1}(t), \lambda) / x'(x^{-1}(t)) = f(\lambda)$, and there remains to determine the functions f and L. We do this by verifying when, with these functions, the integral is indeed independent of λ:

$$\int_0^{L(\lambda)} f(\lambda) e^{2\pi i \alpha t/\lambda} dt = f(\lambda) \frac{e^{2\pi i \alpha L(\lambda)/\lambda} - 1}{2\pi i \alpha/\lambda} \tag{26}$$

which is indeed independent of λ only if $L(\lambda) = L_0 \lambda/\lambda_0$ (we have already assumed that the entire length of the array L_0 is used at the lowest design wavelength λ_0), and if $f(\lambda) = constant'/\lambda$. We summarize these conclusions:

$$w(n, \lambda) = constant' \frac{x'(n)}{\lambda} \text{ if } n < x^{-1}(L_0 \lambda/\lambda_0), \tag{27}$$

and should vanish otherwise.

What should the overall design be in light of all this? The first halves of conditions (19) and (27) specify the support of the shading function w. To reconcile them, we must succeed in construction $x(n)$ such that

$$x^{-1}(L_0 \frac{\lambda}{\lambda_0}) = x'^{-1}(x'(N) \frac{\lambda}{\lambda_0}) \tag{28}$$

Using the new variable $\lambda' = L_0 \lambda/\lambda_0$, this equation becomes $x^{-1}(\lambda') = x'^{-1}(x'(N)\lambda' L_0)$. If we now write ν for $x^{-1}(\lambda')$ (so that λ' becomes $x(\nu)$) then apply x' to both sides of the equation, we obtain

$$\frac{dx}{d\nu} = \frac{x'(N)}{L_0} x \tag{29}$$

a differential equation which is solved by $x(\nu) = Ae^{x'(N)\nu/L_0}$, in which A is a constant that can be determined by the condition $x(N) = L_0$. Putting it all together, The solution is

$$x_n = L_0 e^{-\frac{x'(N)}{L_0}(N-n)} \tag{30}$$

This equation must hold at least for all n $(= x^{-1}(L_0 \lambda/\lambda_0))$ that are within the interval $[\lambda_1, \lambda_0]$ where λ_1 is the wavelength corresponding to the highest design frequency. Also note for future use that (30) implies that, over its range of validity, $x(x'^{-1}(d)) = L_0 d/x'(N)$.

With formula (30) for x_n, condition (22) governing the sidelobe levels becomes

$$w(n,\lambda)^2 = constant\ 1/L_0 \tag{31}$$

for all n in the range $x'^{-1}(\lambda/2) < n < x'^{-1}(x'(N)\lambda/\lambda_0)$ and all λ in $[\lambda_1, \lambda_0]$. It is important to note that in order to validate the use we just made of formula (30), we must require that that formula hold for all n in the interval $[x'^{-1}(\lambda/2), x'^{-1}(x'(N)\lambda/\lambda_0)]$ over all λ in the design band. Since x' is increasing, this means that (30) must hold for all n in the interval $[x'^{-1}(\lambda_1/2), N]$. As to our amplitude shading function w, we may as well normalize it by setting the constant in (31) equal to L_0, so that $w(n,\lambda) = 1$ over the interval above.

To specify w over the entire domain of interest, we turn to the last condition unaccounted for yet, namely the second half of (27), namely that $w(n,\lambda) = constant' x'(n)/\lambda$ if $n < x^{-1}(L_0\lambda/\lambda_0)$. Unfortunately, to maintain a low sidelobe level, we have already adopted the condition $w(n,\lambda) = 1$ for $x^{-1}(\lambda/2) < n < x^{-1}(L_0\lambda/\lambda_0)$. So we will simply choose the constant in (27) in such as way as to make w continuous at the transition point $n = x^{-1}(\lambda/2)$, i.e. we set $constant' = 2$, and set

$$w(n,\lambda) = \frac{2}{\lambda} x'(n) \text{ for } n < x'^{-1}(\frac{\lambda}{2}) \tag{32}$$

To obtain a complete design, we still need to specify x_n for all $n = 1 ..., N$. So far we know that (30) must hold for all n in the intersection I of the intervals $[x^{-1}(L_0\lambda_1/\lambda_0), N]$ and $[x'^{-1}(\lambda_1/2), N]$. Therefore we still need to determine $x'(N)$ and the values of x_n for n outside I.

Assuming, without loss of generality, that $x'(N) > \lambda_0/2$ (i.e. that the largest hydrophone spacing is at least as big as half the largest wavelength in the problem), our range I is in fact the interval $I = [x'^{-1}(\lambda_1/2), N]$. Thus, (30) specifies the element positions from the widest-spaced down to the ones spaced at half the shortest design wavelength. The remaining elements (whose positions are yet to be specified) must therefore be spaced at least as closely as half the wavelength at any frequency in the design band. For efficiency, it is only natural to minimize the number of additional elements available to fill out the array, and therefore require that they be spaced no closer than the half-minimum-wavelength. So we set

$$x_n = n\frac{\lambda_1}{2} \text{ for } n \leq x^{-1}(\frac{L_0}{x'(N)}\frac{\lambda_1}{2}) = N - \frac{L_0}{x'(N)} \log(\frac{x'(N)}{\lambda_1/2}) \tag{33}$$

In order that x remain continuous, we must finally reconcile this last equation with equation (30) at the intersection point of their ranges. This reduces to imposing the final additional condition

$$\frac{\lambda_1}{2}\left(N - \frac{L_0}{x'(N)} \log(\frac{x'(N)}{\lambda_1/2})\right) = \frac{\lambda_1}{2}\frac{L_0}{x'(N)} \tag{34}$$

Putting all these requirements together, we end up with the following equations for the element positions and the shading coefficients (here we call d the largest spacing in the array):

$$d \geq \lambda_0/2 \tag{35}$$

$$L_0 = \frac{dN}{1 + \log(2d/\lambda_1)}) \tag{36}$$

$$x_n = n\lambda_1/2 \text{for} n \leq N - \frac{L_0}{d}\log(2d/\lambda_1) \tag{37}$$

$$x_n = L_0 e^{-(N-n)d/L_0} \text{ for } n \geq N - \frac{L_0}{d}\log(2d/\lambda_1) \tag{38}$$

$$w_n(\lambda) = 2x'(n)/\lambda \text{ for } n \leq N - \frac{L_0}{d}\log(2d/\lambda) \tag{39}$$

$$w_n(\lambda) = 1 \text{ for } N - \frac{L_0}{d}\log(2d/\lambda) \leq n \leq N - \frac{L_0}{d}\log(\lambda_0/\lambda) \tag{40}$$

$$w_n(\lambda) = 0 \text{ otherwise.} \tag{41}$$

Practically, this implies that the filtering we apply to the Fourier coefficients $\hat{r}_n(f)$ at frequency f of the signal r_n at the n^{th} hydrophone is

$$\sum_n (A_n(f) + \int B_n(f))\hat{r}_n(f)e^{2\pi i f \beta x_n/C} \tag{42}$$

with x_n as above, $A_n(f) = \lambda_1/C$ if $n \leq N - \frac{L_0}{d}\log(2d/\lambda_1)$, $A_n(f) = (2d/C)e^{(n-N)dL_0}$ if $N - \frac{L_0}{d}\log(2d/\lambda_1) \leq N - \frac{L_0}{d}\log(2df/C)$, and 0 otherwise, while $B_n(f) = 1$ if $N - \frac{L_0}{d}\log(2df/C) \leq n \leq N - \frac{L_0}{d}\log(\lambda_0 f/C)$, and 0 otherwise, and β is the sine of the bearing angle being scanned.

As the simulations that we have implemented show, this method should prove very useful in concentrating the energy of a signal spread over a wide band and thus enhancing its detectability. Indeed, the method singles out the optimal set of time delays (defined by the angle of peak gain) to use in adding up coherently the signals arriving at the different elements of our array.

We have previously shown the benefits of an array in signal classification (see figures 13 and 14). In the next example we show how one can determine the direction of a broad band signal, even when a stronger signal is present. In Figure 15 we show an example with two sources, one at $-45°$ and one at $+30°$. The array does an excellent job of locating the direction. Knowledge of the direction is important for optimum utilization of the array for classifying signals. Indeed, the method singles out the optimal set of time delays (defined by the angle of peak gain) to use in adding up coherently the signals arriving at the different elements of our array.

6. NONLINEAR TRACKING

In this case, we are assuming that a "parametrized" signal template time-series has been identified. By "parametrized," we mean that this template, which we have as a model for the signal that we are trying to detect and localize in the received time series, may depend on several parameters such as target aspect, speed, range, etc. We also assume that, while the dynamics of this template signal are not necessarily linear, they are not chaotic: in other words, we can expect to estimate in a robust fashion the values of the parameters on which

it depends, and, indeed, locate the exact version of our signal within the received time series. For convenience, let us group all our parameters into a "state" vector X, and assume that we are looking for the signal $S(t;X)$ within the received time series $r(t)$.

So far, we have not accounted for the (highly non-linear) propagation effects through the ocean lens. For simplicity, we will assume for now that the parameter vector X consists only of the coordinates of the source producing the time series. Our problem in this case is to look for each of the n signals

$$S_n(t;X) = \int G(X,t_0;X_n,t)\,S(t_0;X)\,dt_0 \tag{43}$$

within the corresponding signal $r_n(t)$ received at the n^{th} hydrophone, where G is the Green's function for the medium, and X_n represents the coordinates of the n^{th} array element. Rather than try to detect S (and estimate its parameters X) based on a single "look," an approach whose success is assured only when one assumes an unrealistically high signal-to-noise ratio, we propose to combine several looks together in order to make our detection. In short, we set out to "track" S (and X) in order to determine if it is indeed there, i.e. for what value of X , if any, can we identify our template within the received time series.

The optimal filtering solution to this problem is described by the formalism of the Zakai equation ([6]). The solution to this equation is essentially the (infinite-dimensional) non-linear equivalent of the linear Kalman filter. To set up the Zakai equation, and then try to find its solution, one starts by describing the expected time evolution of the vector one is trying to estimate, X in our case. One then describes the dependence of the data received, the r_n's in our case, on this state vector - that is achieved by our equation (43), once we can compute the medium's Green's function. For simplicity, we shall assume very simple dynamics for our variables, and proceed to write down directly the solution to the Zakai equation: first, we change the assumption about X a little, and assume that our state vector consists of the coordinates of the source as well as its velocity vector, i.e. $X = (x^{(1)}, x^{(2)}, x^{(3)}, v^{(1)}, v^{(2)}, v^{(3)})$, then we assume that the dynamics of the velocity vector have a deterministic part that assumes that the velocity is piecewise constant, and a stochastic part that assumes that the discrete velocity changes occur at Poisson-distributed times with rate μ. This measn that our source moves in straight line segments, changing its course at a sequence of times $\{t_1, t_2, t_3...\}$ such that the quantities $t_1, t_2 - t_1, t_3 - t_2$, ... are Poisson distributed with mean $1/\mu$, and such that the new velocities at each course change are distributed according to a pre-specified density function, call it g. The formalism of the Zakai equation allows us to determine the conditional density function $\rho(t, Y$ given $r_n(\tau)$ for all $\tau \leq t)$, the function of (t, Y) which describes the likelihood that X at time t be close to the value Y, given all the past received observations $r_n(\tau)$.

Under these assumptions, we have obtained a recursive algorithm for updating ρ as the data is received. Indeed, if the noise in the received signal is 0-mean additive Gaussina noise with r.m.s. level σ, ρ can be computed using

$$\rho(t,x,v) = \rho(t_0, x-(t-t_0)v, v)\,e^{-\frac{1}{\sigma^2}\int_{t_0}^{t}\sum_n [S_n(s,x-(t-s)v,v)^2 + 2u_n(s)\,v\nabla S_n(s,x-(t-s)v,v)+\mu]\,ds} \tag{44}$$

$$+\mu \int_{t_0}^{t} [\int \rho(t_0, x-(t-t_0)v, v')g(v-v')dv']e^{-\frac{1}{\sigma^2}\int_{t_0}^{t}\sum_n [S_n(\tau,x-(t-\tau)v,v)^2+2u_n(\tau)\,v\nabla S_n(\tau,x-(t-\tau)v,v)+\mu]\,d\tau} ds \tag{45}$$

where $u_n(s) = \int_0^s r_n(\tau)d\tau$ is the sum of all data received at the n^{th} element up to time s.

We set out to study the moments of ρ and their evolution in time in order to determine the kind of localization one can expect to achieve efficiently. The preliminary simulations we had conducted started by assuming a medium with given visco-elastic propagation properties, then computing the associated Green's function for a pure tone, thus enabling us to compute the templates S_n for any array configuration.

The first new aspect of the problem we decided to study was the robustness of this approach when one does not necessarily know the exact parameters governing the propagation in the medium at hand. Indeed, one cannot expect to know the compressional and shear velocity profiles with all their fluctuations throughout the medium, nor can one expect to compute an exact Green's function, even if the propagation parameters were known exactly. In particular, starting with a uniform ρ over a realistic domain for X, the first question is to determine how soon the variance of ρ, i.e. the mean square uncertainty in the estimate of X, would converge to an acceptably small value. We started by doing this for a pure tone (the case of a linear signal), in a two-layer medium allowing only compressional wave propagation. The discouraging result is that the method fails if the relative discrepancy between the modeled velocity and the actual propagation speed is greater than 0.4%, if the relative difference between the modeled and actual layer depth is greater than .1%, or if the relative error in the density of the half-infinite layer is greater than 10%. Not only does the second moment fail to converge, but the algorithm actually ends up "zeroing" ρ because the increasingly negative exponents in (44) impose a uniform rapid decay rate on the density function. In essence, the "ambiguity function" for this particular approach is too narrowly supported, and this model is too sensitive to randomness in the medium.

The conclusion is that for this approach to work, one must incorporate into one's vector of "state variables" the parameters that describe the medium, and one must incorporate the effect of their fluctuations on the signal as it propagates from the source to the receivers. This is considerably more complicated than what we set out to study: such an endeavor would lie beyond the scope of this effort.

References

[1] Chow, Y.L., *On grating plateaux of nonuniformly spaced arrays*, I.E.E.E. Trans. on Antennas and Propagation, Vol. 13, March 1965, pp. 208-215.

[2] Haddad, Z.S. and B.E. Parkins, *Designing Doppler signals to overcome transmission nonlinearities* , I.E.E.E. Trans. on Signal Processing, Vol. 40, October 1992, pp. 2459-2464.

[3] Ishimaru, A. *Theory of unequally spaced arrays*, I.E.E.E. Trans. on Antennas and Propagation, Vol. 10, November 1962, pp. 691-702.

[4] Ishimaru, A. and Y.S. Chen, *Thinning and broadbanding antenna arrays by unequal spacing*, I.E.E.E. Trans. on Antennas and Propagation, Vol. 13, January 1965, pp. 34-42.

[5] *Numerical Recipes*, W. Press *et al*, Cambridge University Press, 1992.

[6] Zakai, M., *On the optimal filtering of diffusion processes*, Z. Wahrsch. verw. Geb., # 11, 1969, pp.230-243.

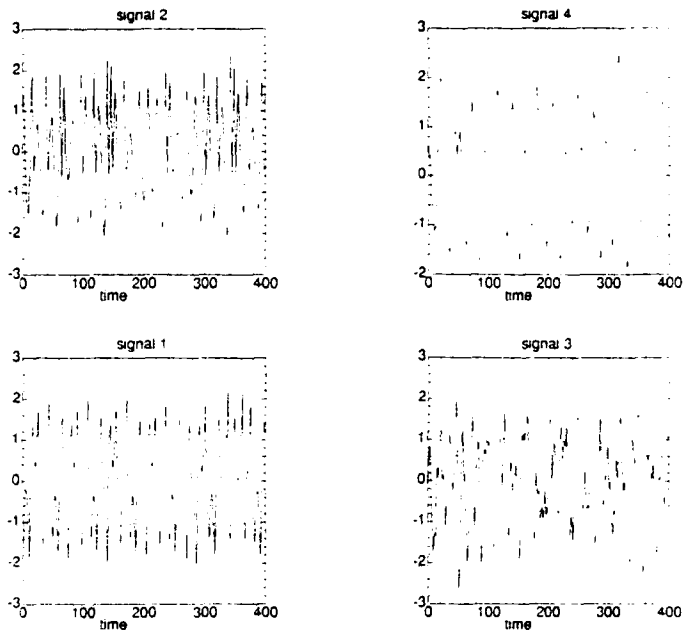

Figure 1. Time traces of four of the signals used in this demonstration. Signal 3 is random gaussian noise with the same spectrum as one of the other signals.

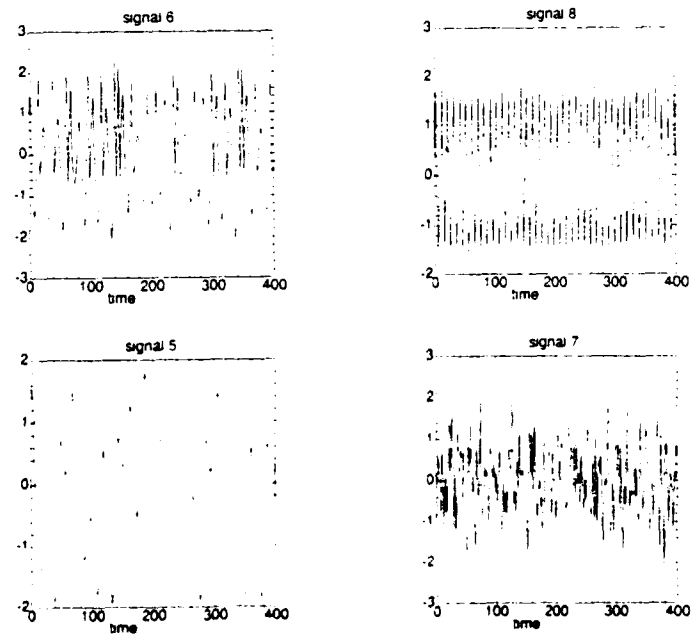

Figure 2. Time traces of four of the signals used in this demonstration. Signal 7 is random gaussian noise with the same spectrum as one of the other signals.

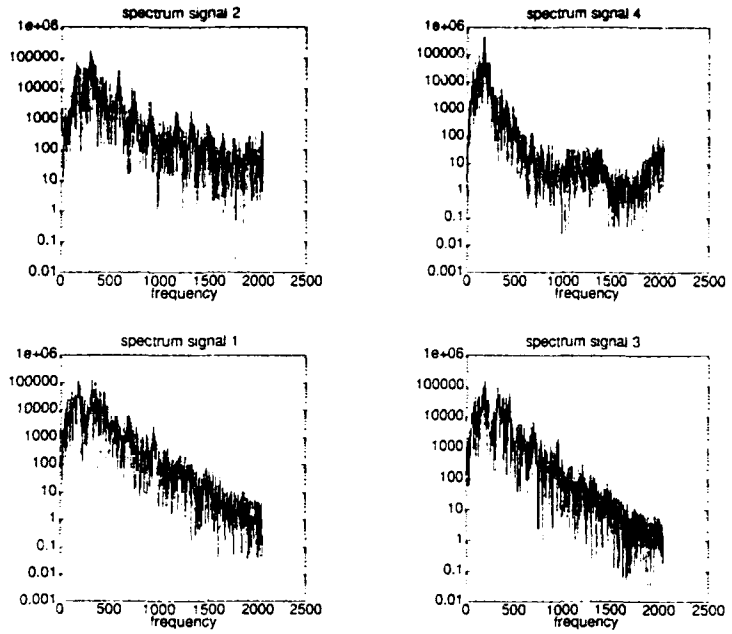

Figure 3. The corresponding power spectra for the signals shown in Figure 1.

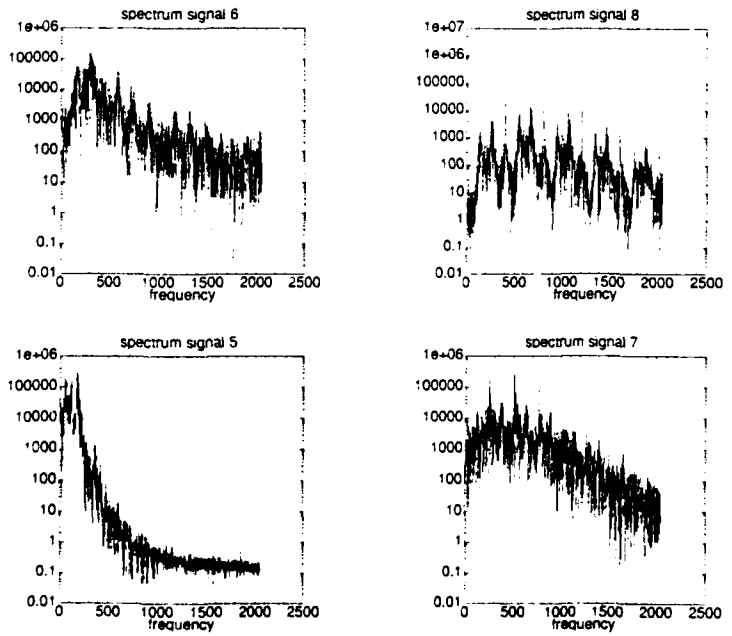

Figure 4. The corresponding power spectra for the signals shown in Figure 2.

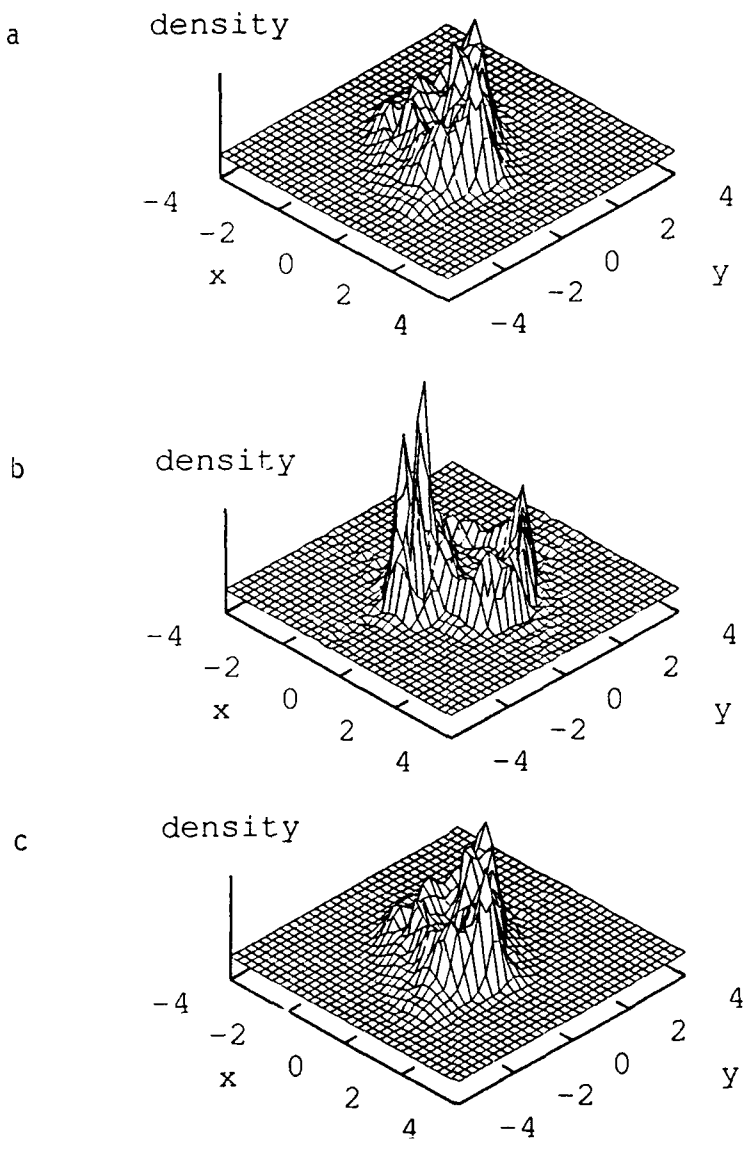

Figure 5. The density, $\rho(x, y)$, for a particular signal is shown in Figure 5. The density from another sample of the signal is shown in 5c, and the density from a different signal is shown in 5b. These latter two densities are examples of proposed candidates taken from a library of possible densities.

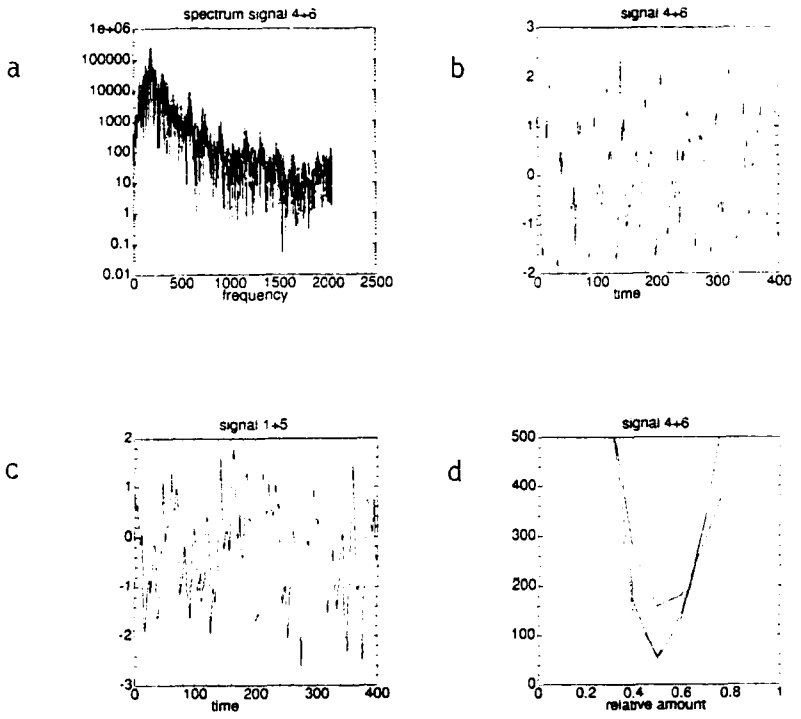

Figure 6. (a) is the power spectrum of the signal shown in 6b; (b) is the sum of signals 4 + 6 with the relative strength parameters a and b given in the text; (c) Different pair of signals that was being compared with signals 4 + 6; (d) χ^2 is a figure of merit for how good the fit is, smaller χ^2 being better. The horizontal axis is a relative amount of the two signals being combined. The two curves that nearly touch are two samples of signals 4 + 6. The upper curve is for signal 6 + another unshown signal, and the χ^2 for signals 1 + 5 would be off the graph.

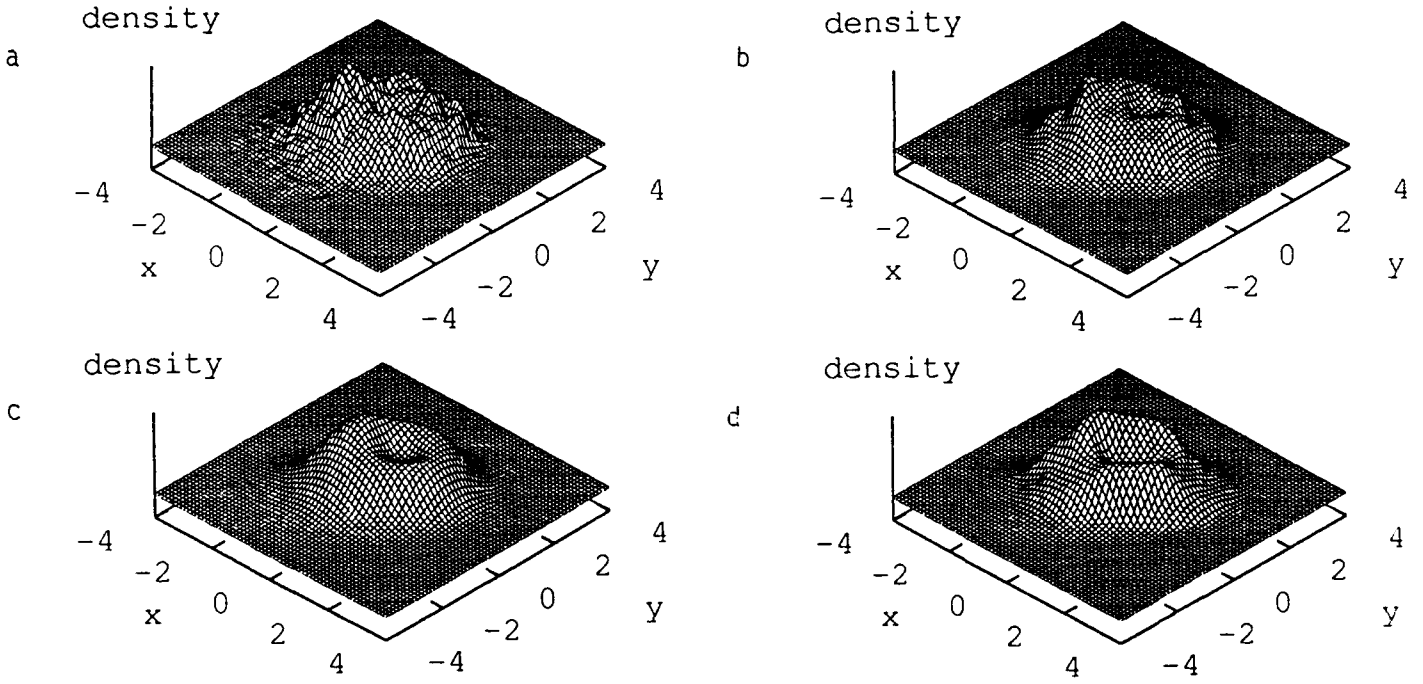

Figure 7. Figure 7a is the density, $\rho(x, y)$, for $S = .37S_1 + .63S_2$ with S_1 and S_2 being two of the signals described above. Figure 7b is the density constructed from $\hat{\rho}(\kappa) = \hat{\rho}_1(.37\kappa)\hat{\rho}_2(.63\kappa)$ and would be identical to Figure 7a for an infinitely long time series. Figures 7c and 7d are constructed similarly to Figure 7b, but the signal used for S_1 is incorrect.

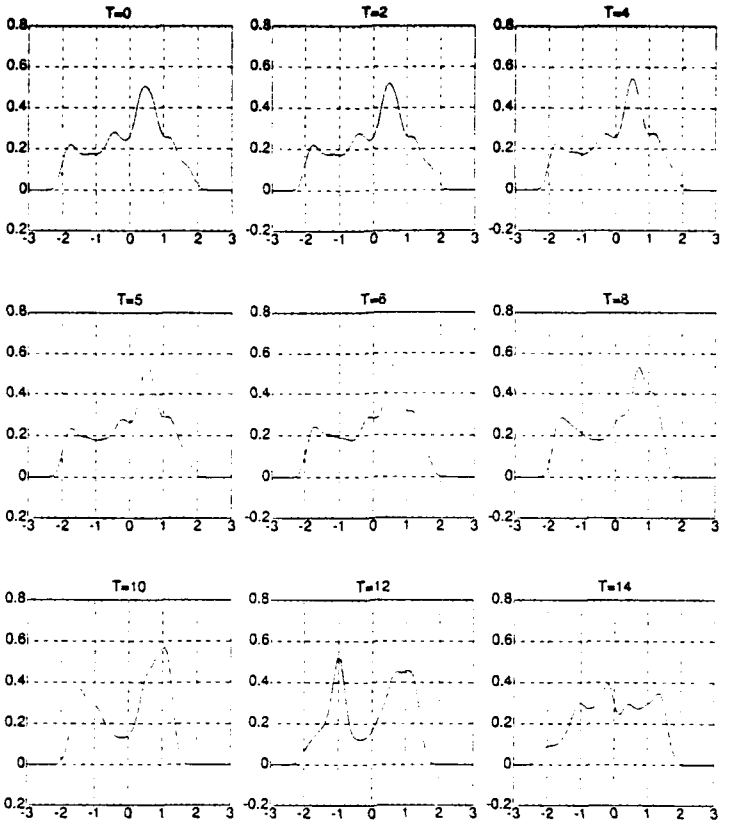

Figure 8. One-dimensional density for one of the signals with $S_{sum}(t) = S(t) + S(t+T)$ and various time lags, T. The vertical axis is the density and the horizontal axis is the signal amplitude.

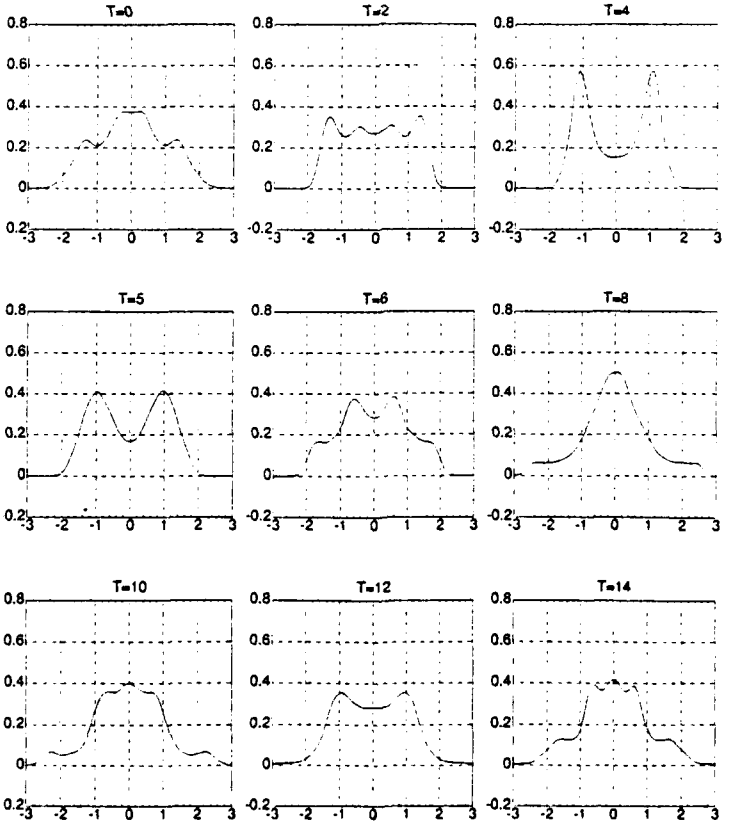

Figure 9. Same as Figure 8 for another signal.

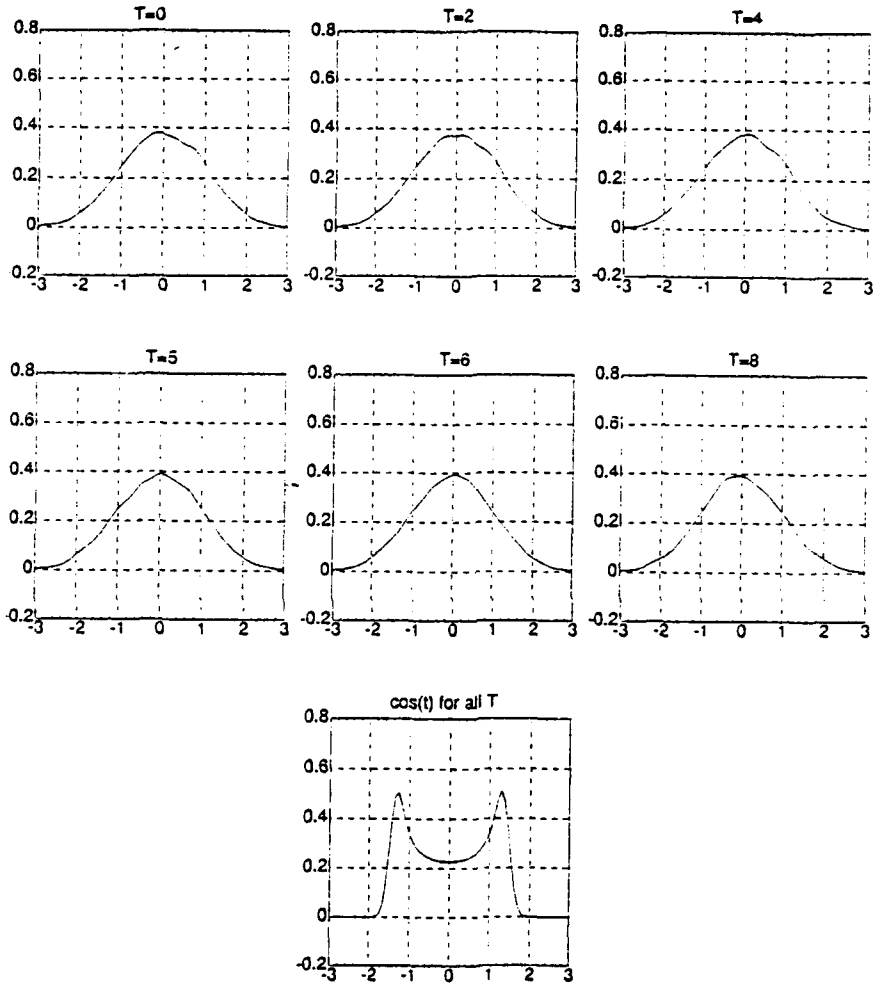

Figure 10. The signal has the same spectrum as the signal in Figure 9, but the Fourier phases were randomized to produce a noise signal. The single figure on the last row is for $S(t) = \cos(t)$. The densities for $\cos(t)$ are independent of the time lag, T. The other densities have very little dependence on T.

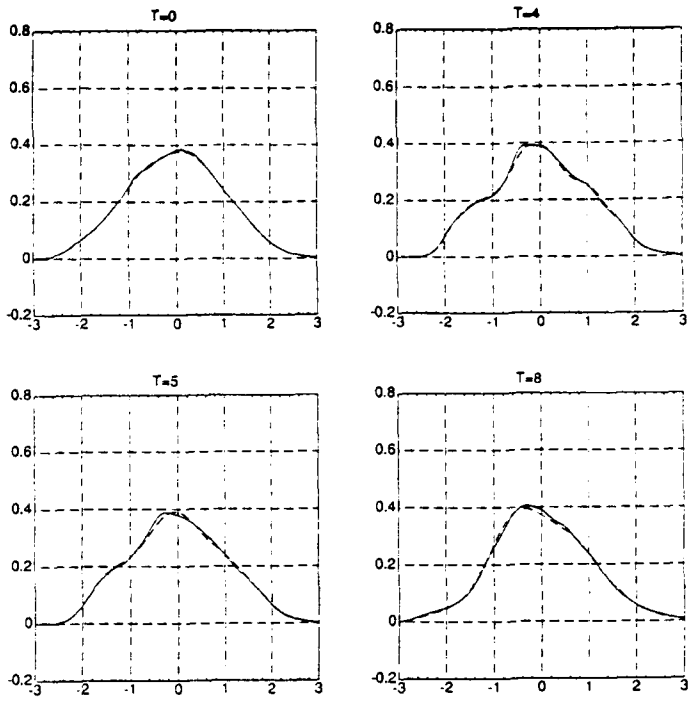

Figure 11. The solid line is the density constructed from $S(t) = S_1(t) + S_2(t)$ where S_1 and S_2 are two of our standard signals. The dashed line is the fit obtained by calculating the convolution of the densities for S_1 and S_2. The difference is due to the limited length of the signal.

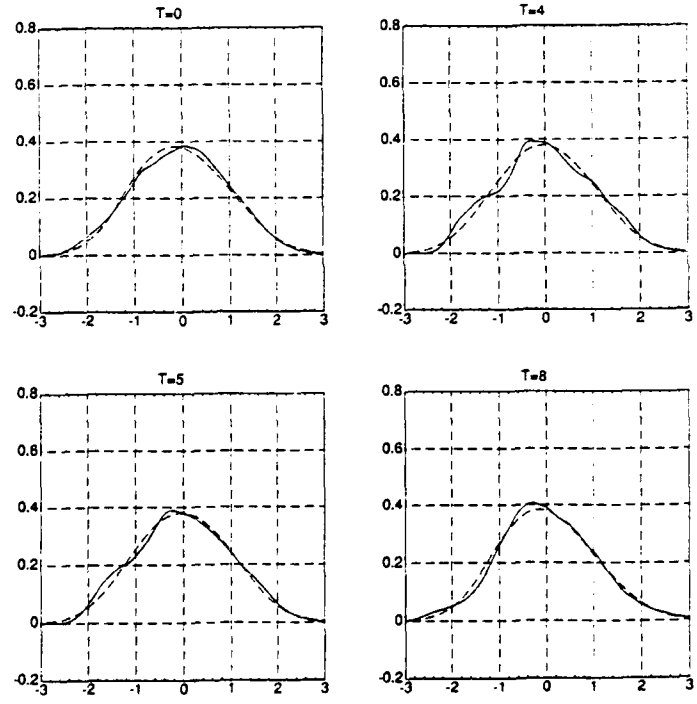

Figure 12. The solid line is the same as that in Figure 11. The dashed line comes from replacing S_1 by another signal (see text) and then calculating the convolution of its density with ρ_2, the density for signal S_2.

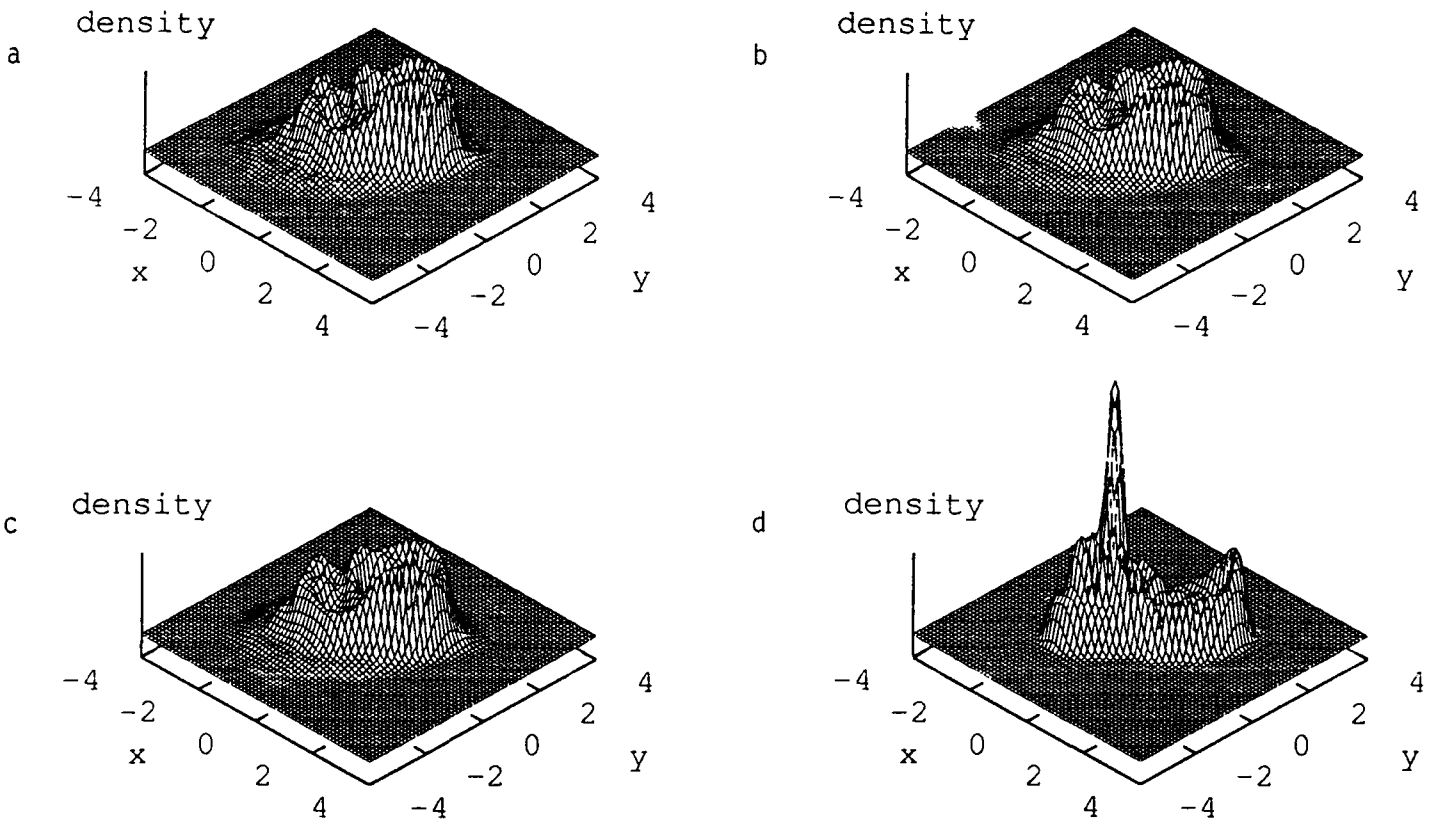

Figure 13. The signal is $S(t) = S_1(t) + 10S_2(t)$ and Figure 13a shows the density for S. Figure 13b is the density obtained from $\hat{\rho}\ \vec{\kappa} = \hat{\rho}_1\left(\frac{\vec{\kappa}}{10}\right)\hat{\rho}_2\ \vec{\kappa}$. To obtain Figure 13c, we construct a trial $\hat{\rho}$ from a signal other than S, and then Fourier transform it. Figure 13d is similar except that a signal other than S_2 is used. We see that with a signal this weak we can not identify it with a simple receiver and a single time lag, although the strong signal is readily identifiable.

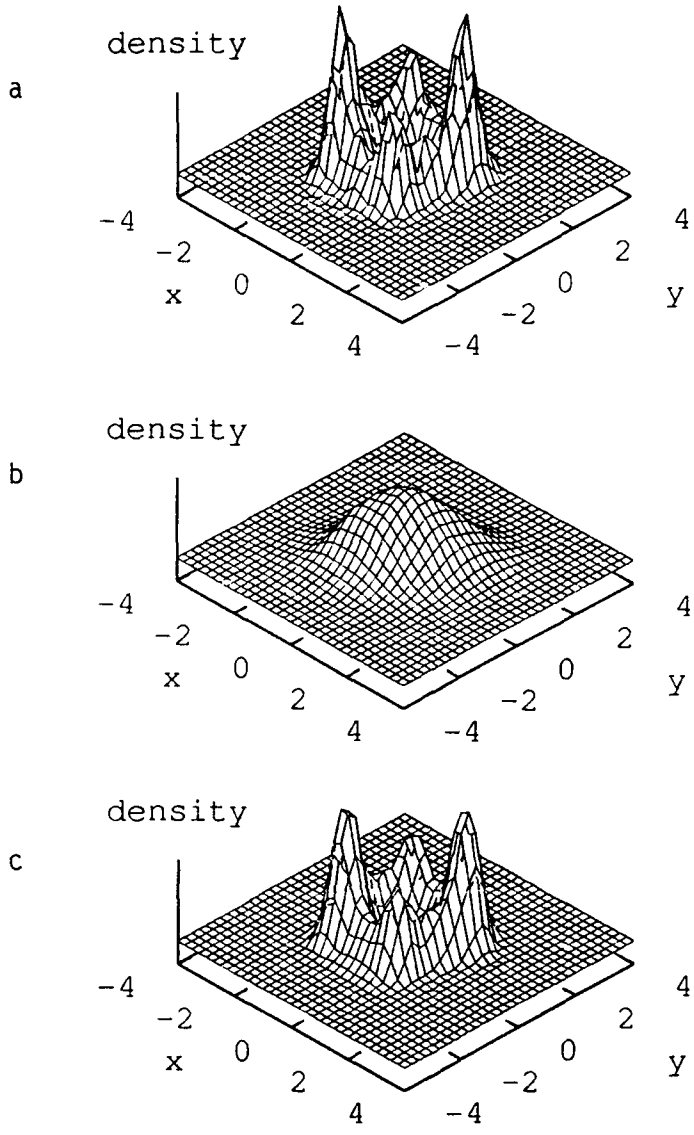

Figure 14. Two signals were directed at an array with an angle of 45° between them. Signal 1 had an amplitude 10 times signal 2, and the array had 100 elements. The array was pointed at the weak signal. The two-dimensional density of the signal is shown in Figure 14b. Two sample hypotheses are shown in Figures 2 and 3. These two signals happen to have identical power spectra. Clearly Figure 14c can be easily selected. Without the array and with only one time lag used, we could not have identified the correct component.

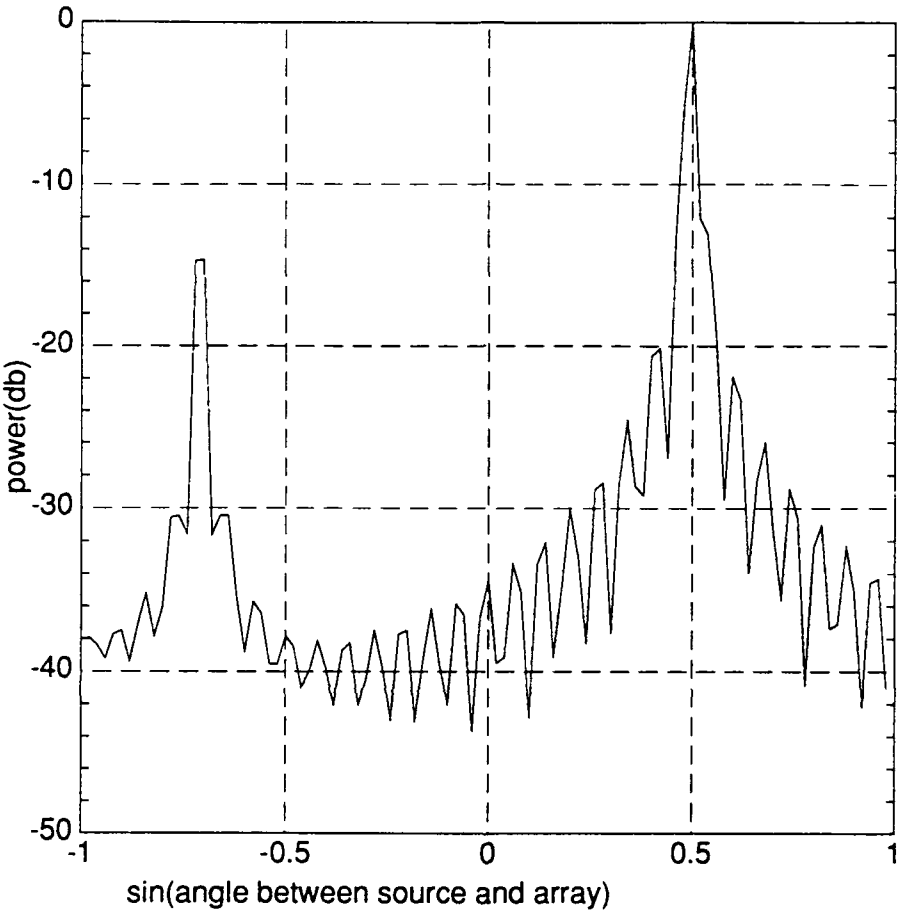

Figure 15. Two signals were simulated, one at -45° and one at +30° with respect to the array orientation. The phase of the array was varied and the return is plotted versus the sin of the direction angle. The source directions are clearly picked out.